MPhil (Mathematical Finance)

STA5088H: Thesis

Stock Price Fragility in an Emerging Market

Author:
JL Nairac
nrcjea001@uct.ac.za

Supervisor:
Dieter Hendricks
dieter.hendricks@bnpparibascadiz.com

May 19, 2013

Plagiarism Declaration

I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

Each significant contribution to, and quotation in, this honours project from the work of others has been attributed, cited and referenced.

This project is my own work.

I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Signature: ______________________________

Acknowledgements

I would like to thank the following people for their valuable contribution to my project:

- To Dieter Hendricks, for his guidance, supervision and helping to get through all those data issues.
- To my family and friends for their support.
- To UCT for providing me with all the knowledge and patience to tackle projects like this one.

Abstract

This research project examines stock price fragility, a measure developed by Greenwood and Thesmar (2011), which serves as a proxy for non-fundamental risk i.e. it aims to isolate the drivers of stock price volatility beyond traditional fundamental drivers, in particular examining the impact of concentrated stock ownership and correlated liquidity shocks on price volatility. Here, the measure is applied to the South African financial market. Subject to data complications, it is nevertheless shown that stock price fragility is a significant predictor of total return volatility owing to the ownership structure of South African funds, even when controlling for endogeneity, autocorrelation and heteroskedasticity in the model. Also developed by Greenwood and Thesmar (2011), the forecast of the covariance and beta of returns, by the co-fragility and the fragility beta measures respectively, is explored. Here, although significance of these coefficients cannot be inferred, it is suggested that the ownership structure of funds has impact on these forecasts. Finally, Greenwood and Thesmar (2011) explore the sensitivity of stock price fragility to total return volatility through the impact of arbitrageurs on stock prices. In this research project, the impact of arbitrageurs is investigated and shown to be significant in the South African market context.

Table of Contents

List of Illustrations

Figures

Tables

1

Introduction

In this chapter, the research topic is introduced by providing a background to the topic, a review of the existing literature as well as the major objectives of this research project. Chapter 2 provides the methodology behind the project, guided by the methodology used in Greenwood and Thesmar (2011). Chapter 2 is specifically concerned with an analysis of the data as well as the theoretical construction and the empirical methods for stock price fragility, co-fragility, the fragility beta and the impact of arbitrage. Chapter 3 provides and analyses the results obtained through the processing of the methodology. The results include those from the implementation of stock price fragility, co-fragility, the fragility beta and the impact of arbitrage. Chapter 4 provides the salient conclusions to this research. Finally, after the chapter of references, chapter 6, the appendix, provides additional information which may be useful to the reader.

1.1 Background & Overview

Adequately foreseeing the risk involved when investing in a stock is a problem that all investors face. This is, for the most part, due to the fact that its risk is difficult to quantify. One method of quantifying risk is to break down into its several components and summing them analogous to arbitrage pricing. The problem here then is to determine what makes up each component. Some components of a stocks risk may be driven by the stocks associated news or fundamentals e.g. book-to-market ratio. Another component may be a proportion of the risks time dependence. These components of risk have of course been tackled by many over the years and are yet to be perfectly quantified. And yet, possibly even more complex to measure are the remaining components of risk. These components of risk are known by most investors as unexplained, idiosyncratic, endogenous or non-fundamental risk. Examples this kind of risk include the perception of risk, investor irrationality, individual or idiosyncratic liquidity shocks, etc. This research project is concerned with the latter example.

One of the responsibilities of a portfolio manager is to actively manage the investors portfolio i.e. rebalance the portfolio on a frequent basis in response to fundamental changes in the stocks within the portfolio. A portfolio manager may also be managing a pool of investors money, usually known as a fund. The fund is actively managed but, the manager general has little control over the cash inflows and outflows of the fund this is left to the investors discretion. Thus, it suffices to suggest that, in general, the reasons behind fund flows are non-fundamental and the risk associated with these flows contribute to non-fundamental risk. However, the level of contribution to this component of risk may be very small simply because there may be a large number of investors with uncorrelated liquidity needs in the fund. Put another way, a single investor drawing a small portion from the fund will unlikely have any price impact on the stocks within the fund. Hence, the contribution to non-fundamental risk will be dependent on the ownership concentration of the assets in the fund as well as the volatilities and correlations of fund flows. More broadly, this concept may be applied to the whole universe of stocks and not just the portion of those held by funds.

This research project builds on a paper by Greenwood and Thesmar (2011) in which a measure for non-fundamental risk is formalized. The authors define this measure to be stock price fragility and is derived from the ownership structure of assets as well as the volatilities and correlations of the investors liquidity needs, explained above. Although the measure does not fully capture all forms of non-fundamental risk, it may serve as an adequate proxy for it, making it an interesting project to take on. More specifically, this project attempts to assess stock price fragility, including some of its natural extensions, in the context of the South African financial market. The driving factor behind working in this context is the idea that the liquidity structure of an emerging market is likely to be different to that of a developed one. For instance, investors may be more or less responsive to shifts in liquidity resulting in a more or less pronounced effect of stock price fragility on total return volatility respectively.

1.2 Objectives

The objectives of this project are fulfilled in the following sequence:

1. Initially, the objective is to understand, assimilate and analyse the derivation and practical applications of stock price fragility presented in the paper by Greenwood and Thesmar (2011). Put another way, the aim here is to understand the concept of non-fundamental risk and to contrast stock price fragility against existing attempts at formalizing a measure or proxy for non-fundamental risk.

2. The next objective is to understand how we would apply the methodology followed in Greenwood and Thesmar (2011) in the context of the South African financial market in

order to contribute to existing literature. Moreover, to see whether any useful additions to the methodology can be made. The data pertaining to the South African financial market must then be collected, analysed, sorted and classified to suit the needs of this research.

3. Then, we implement and assess stock price fragility using the data by the methodology applicable in this context. We do the same for co-fragility and the fragility beta, a natural extension to stock price fragility.

4. Finally, we assess the impact of arbitrage on flow-induced trading - another natural extension to stock price fragility.

1.3 Literature Review

This literature review first introduces the concept of non-fundamental risk and its major components. This is followed by a review of the attempts at measuring non-fundamental risk. At the core of this review is the paper by Greenwood and Thesmar (2011), providing one such measure. The authors' methodology is discussed throughout the rest this review with application to South African financial market.

Modern portfolio theory (Markowitz, 1952; Treynor, 1962; Sharpe, 1964; Lintner, 1965) and arbitrage pricing theory (Ross, 1976) are highly influential in what is considered as conventional asset pricing theory. In deriving standard pricing models, both rely on several assumptions including: all investors must have homogenous expectations and all investors must be price takers. Traditional theoretical pricing models thus do not explain asset ownership structure as an explanatory factor for risk and return. This implies that if existing owners of an asset are forced to buy or sell for reasons not determined by fundamental factors, then frictionless trading with counterparties to their transactions would always be possible. Moreover, idiosyncratic shocks are assumed to be uncorrelated with asset risk and return and with the fundamental factors that drive them, resulting in flat individual investor demand curves. In contrast, recent research has shown, with application to the US financial markets, that idiosyncratic liquidity shocks do have price impact (Coval and Stafford, 2007; Frazzini and Lamont, 2008; Lou, 2010). Research has inferred similar results within the context of the South African financial market: Pillay, et al. (2010) found there to be a relationship between funds sizes, containing JSE listed stocks, and their performance. The authors also found that the funds' liquidities are the underlying cause of this relationship.

A review of this literature reveals implicitly that idiosyncratic (or unexplained) shocks play a role in predicting returns and the risks associated with these shocks pose as a factor in predicting the volatility of returns. The risk associated with idiosyncratic shocks may be defined as non-fundamental or endogenous risk. Put another way, endogenous risk refers to the risk from shocks that are generated and amplified within the system (Danielsson and Shin, 2003, p.1). It is driven by information unrelated to fundamentals (news). This form of risk has been shown to contribute more to total return volatility (Kurz, 2000; Danielsson and Shin, 2003) than its counterpart, exogenous risk. This has made the study of non-fundamental risk valuable in forecasting risk and return. Authors such as Greenwood and Thesmar (2011) as well as Cont and Wagalath (2012) have contributed to this study by formalizing a measure of non-fundamental risk. Although both adopt a different approach, they both rely on the nature of liquidity shocks as a basis for constructing this measure. The literature reveals little or no research on forecasting non-fundamental risk in the context of the South African financial markets. Thus, this review primarily focuses on the measure provided by Greenwood and Thesmar (2011), namely: stock price fragility.

Greenwood and Thesmar (2011) define stock price fragility as the expected volatility of non-fundamental demand given an asset's ownership structure. It is expressed as a function of an asset's ownership structure as well as the ex-ante variance-covariance matrix of the liquidity needs faced by its owners. Stock price fragility thus serves as a proxy for non-fundamental risk which may be used as a predictor of total asset return volatility. According to the authors, ownership data may be observed for the majority of assets, however, acquiring information on the liquidity needs of investors is more complicated. Therefore, the authors suggest extracting mutual funds data. That way, ownership structure of the funds may be inferred from the assets within them and the variance-covariance matrix of investors liquidity needs may be ascertained from the flows into and out of those funds. With mutual funds data alone, Greenwood and Thesmar (2011) were able to show that stock price fragility is a significant predictor of total volatility, even when controlling for other determinants of total volatility suggested by existing literature. These additional determinants of volatility aid in addressing the authors concerns of endogeneity and omitted variable bias. These include lagged volatility, lagged skewness, firm size, book-to-market ratio, share turnover, etc. However, intuitively speaking, the structure of return volatility in an emerging market (such as SA) is likely to be different from that of a developed one (such as the US). It is for this reason that reviewing the existing literature regarding predictors of return volatility in SA is necessary.

Historical research on return volatility prediction in the context of the South African equity market is scarce. However, there is some available information in more recent literature. Samouilhan and Shannon (2008) forecasted the volatility of returns on the JSE and found volatility to be dependent on lagged volatility. These finding are supported by Mangani (2008). The authors also show that asymmetric models of volatility are best suited for forecasting. This indicates that lagged skewness may serve as an additional predictor in forecasting SA return volatility. Samouilhan and Shannon (2008) further find significance in implied volatility instead of historical volatility. These findings are supported by Caicedo-Llano and Dionysopoulos (2007). The latter authors also find that both dividend yield and credit spreads are useful in predicting stock market movements. Yartey (2008) found evidence that credit spreads comove with stock market capitalization and since Jefferis and Smith (2005) strongly support market capitalization as a predictor of SA return volatility, multicollinearity may be avoided by replacing credit spreads with market capitalization as a predictor. Bekaert et al. (2007) found that unexpected liquidity shocks are correlated with shocks to the dividend yield. Therefore, one must be wary when regressing both dividend yield and fragility on total return volatility. Jefferis and Smith (2005) also support share turnover as a predictor of volatility. Moreover, Bae, et al. (2002) present findings that share turnover, firm size as well as investibility (the degree to which a stock can be foreign-owned) are useful in predicting return volatility in the context of emerging markets. One may avoid investibility as a predictor here because little research has been done on this measure.

Greenwood and Thesmar (2011) use a Fama-Macbeth first-stage forecasting regression (Fama and Macbeth, 1973) i.e. the authors' regression procedure is concerned with running a time-series regression for each security. Then, the mean of the regression coefficients is taken and used to test the null hypothesis. This procedure accounts for the fact that numerous stocks each need to be regressed over different time points. By doing so, the regressions produce standard errors that have been corrected for cross-sectional correlation (however, not for autocorrelation). Additionally, to control for endogeneity in the regressions, the authors also perform panel regressions, used for data with two or more dimensions, with firm fixed effects (see Gormley and Matsa, 2012), which refer to effects which are specific to a firm as opposed to effects which are common to all firms i.e. the observations pertaining to the firm effects are treated as non-random (they do not vary across time). For this procedure, a regression is run whereby the time-varying variables are demeaned and the mean of the fixed effects are taken. These adjusted variables are then used in an OLS regression. Greenwood and Thesmar (2011) also test stock price fragility as a predictor of excess volatility. To do this, the authors implement a one-, three- and four-factor model (Fama and French, 1992; 1993; Carhart, 1997). These factors are the market risk premium, HML (high minus low book-to-market ratio risk), SMB (small minus big market capitalization risk) and MOM (the momentum risk component: prior month winners minus prior month losers). A review of the existing literature reveals that the Fama and French Factor model (1993) works well in the context of the South African financial market. For example, Basiewicz and Auret (2010) justify this model on JSE returns, with the implementation of the Newey-West estimator (1987), see later.

Greenwood and Thesmar (2011) go on to explore two natural extensions of stock price fragility. The first involves constructing co-fragility the covariance between the non-fundamental demands for two assets given their ownership structures and the fragility beta the non-fundamental sensitivity of an asset relative to its benchmark portfolio. With this extension, the idea is that stocks comove if the owners have correlated trading needs. According to the authors' findings, empirical analysis of these extensions validates that they adequately predict comovement. Their findings contribute to research such as Anton and Polk (2010) who show that stocks that are owned in common have correlated equity returns. Koch et al. (2010) show that correlated trading among investors leads to correlated movements in liquidity which may be largely attributed to stocks with high mutual fund ownership. There is also stronger support for these findings in the context of emerging markets (Qin, 2007).

Like the regressions of total return volatility on fragility, Greenwood and Thesmar (2011) run regressions of covariance of returns on co-fragility as well as return betas (including HML, SMB and market betas) on fragility betas, controlled by a host of other time-dependent predictors. The authors suggest controls such as firm size, book-to-market ratios and industry-specific factors justified by the existing literature in developed markets. In the context of the South African market, there is very little literature on predictors of the covariance of returns. Most notably, Mhlanga (2008) finds that industry-specific factors are significant predictors. With regards to return betas, Van Rensburg and Robertson (2003) find that market capitalization is a significant predictor.

The second extension is concerned with determining under which scenarios stock price fragility serves as a better predictor of total asset return volatility. Greenwood and Thesmar (2011) do so by additionally assessing the order imbalances from other groups of investors, including hedge funds and the active trades of US mutual funds. Consistent with intuition and asset pricing theory, the authors show that stock price fragility better predicts total volatility when arbitrageurs trade with the liquidity shocks of other investors. Conversely, the relationship between stock price fragility and total volatility is weakened when arbitrageurs trade against the liquidity shocks of other investors. The authors findings contribute to research such as Brunnermeier and Nagel (2004). For this research project, we primarily focus on the order imbalances driven by the active trades of mutual funds, potentially contributing to research in the emerging markets such as Jotikasthira et al. (2009), who show that there is a significant relationship between mutual fund flows and portfolio reallocations across 25 emerging markets.

Having formalized stock price fragility, Greenwood and Thesmar (2011) illustrate its efficacy as a proxy for non-fundamental risk. However, in order for it to establish itself as a widely-accepted measure for non-fundamental risk, more investigation is required on the measure with application to various contexts and since it is a newly developed measure, there has been very little research on the measure thus far. Though, notably, there are two research papers that apply the measure. Firstly, Lin (2011) investigates the relationship between investor sentiment and stock price fragility, specifically assessing whether funding constraints impact stock market liquidity and how this interacts with investor sentiment. Secondly, Lin et al. (2011) use stock price fragility to investigate how the quality of country governance impacts the volatility of stock market returns. The authors findings revealed that stocks with bad country governance display higher total volatility, idiosyncratic volatility, skewness and lower liquidity.

Now, to hopefully add to the research, the purpose of this project is to assess stock price fragility in the context of an emerging market. With this context in mind, two hypotheses are made. Firstly, based on the findings of Qin (2007), who suggests that because of the higher liquidity risk in emerging markets there is a stronger relationship between it and stock market volatility, and based on the findings of Kaminsky et al. (2001) as well as Bekaert et al. (2007), who show that funds are more volatile in emerging markets owing to the large inflows and outflows of the funds, it is hypothesized that there is a more pronounced relationship between stock price fragility and total return volatility. Secondly, although Greenwood and Thesmar (2011) adequately mitigate concerns of endogeneity and omitted variable bias by adding a host of control variables into the regressions, it is hypothesized that there are more pronounced biases because existing literature on emerging market control variables is not as prevalent as that of developed markets. Furthermore, the findings of Jotikasthira et al. (2009) suggest that in 25 emerging markets inflows and outflows of funds lead to significant portfolio reallocations which in turn have price impact, discussed above. This may indicate stronger correlations between the error terms and the fragility regressors as well as heteroskedasticity in the error terms. Thus, this project seeks to further address these concerns by assessing heteroskedasticity and autocorrelation in the error terms. An additional regression is implemented which produces Newey-West-adjusted Fama-Macbeth standard errors for the regression coefficients. The Newey-West estimator (Newey and West, 1987) is used to overcome autocorrelation and heteroskedasticity in the error terms, ensuring robustness of the models. The procedure begins by taking the Fama-Macbeth estimated coefficients. Then, by a given lag-length (based on the level of autocorrelation) the Newey-West-adjusted standard errors of the regression coefficients are produced.

2

Methodology

For this research project, the methodology followed by Greenwood and Thesmar (2011) is more or less replicated in the context of the South African financial market. Some useful additions have also been made. This methodology is processed through Matlab and Excel.

This chapter may be broken down into five sections. The first section is concerned with the data obtained, the issues that arose with the data and the respective assumptions that had to be made when implementing the methodology. The second section is concerned with theoretically defining stock price fragility and the empirical techniques involved to implement and assess the measure. The third and fourth sections cover the first extension to Greenwood and Thesmar (2011), namely co-fragility and the fragility beta respectively, both theoretically and empirically. The last section covers the second extension to Greenwood and Thesmar (2011) i.e. the impact of arbitrage on the strength of the relationship between fragility and total return volatility.

2.1 Data

Greenwood and Thesmar (2011) formalize the theory for stock price fragility such that it may encompass all owners and stocks. However, the authors could only extract a subset of the whole universe of stocks because not all the flow-induced trades of owners are reported. The authors relied on mutual funds data alone to apply their theory. In the same light, we rely on a small subset containing the returns and the size of 163 funds per month from August 2002 to August 2012. This data set was obtained from Morningstar and, unfortunately, a larger data set could not be acquired from any other source. The size of the data set is far too small to provide reliable results. Thus, in an attempt to mitigate the sampling error when constructing the flows of funds, linear interpolation is used to account for the missing dates over the period. Additional flow data is also constructed from proxy funds based on known funds containing stocks with similar industry characteristics. Note that this may cause endogeneity in the regressions. However, the weights of stocks per fund for 2851 funds, over the above mentioned period, was successfully extracted from I-Net. Given that sufficient data on the ownership structure of funds was obtained, we additionally separate the ownership components of fragility for analysis in this project. With regards to the stocks used, we extract data pertaining to the stocks comprising the JSE Top40 index, in the current period as well as prior periods, where ownership data was available. In total we have a subset of 41 stocks (please see their ticker codes in table A.1 of appendix A).

Greenwood and Thesmar (2011) extract quarterly mutual funds data and compute the quarterly variances and covariances of daily, weekly, bi-weekly and monthly returns on each stock i.e. the authors make adjustments to the time dimension of the stock returns data accordingly. In this research project, we extract monthly data of South African funds. However, we only adjust the time dimension of daily stock returns data. Hence, this project has not been extended to assess the impact of fragility over different time horizons. Moreover, given the fact that we deal with a subset of funds to construct fragility and measure it against the total return volatility of each stock, we expect there to be a less significant relationship. It is felt however that our methodology is sufficient to produce inferences that contribute to literature.

Greenwood and Thesmar (2011) use a host of time-varying determinants of volatility in order to control its potentially endogenous relationship with fragility (e.g. high total return volatility may have been attributed to the selection of funds with both high fragility and volatile fundamentals - this leaves us uncertain as to the cause of this high return volatility). This issue may be exacerbated when working with only a subset of the whole universe of stocks and owners (e.g. it may be the case that this subset contains a larger portion of both volatile flows and volatile fundamentals - selection bias). In this research project, we also address this concern by taking the same course of action. Thus, we extract price data, book value and market capitalization from I-Net for this project. Industry-specific factors are also required in the analysis of co-fragility. Thus, for this project, we extract the standard industrial classification (SIC) code pertaining to each stock. SIC codes are an internationally accepted classification index which allows for the classification of each stock into their respective industries. Please see appendix A for an example as well as the tables of SIC codes for each stock in table A.1.

2.2 Stock Price Fragility

2.2.1 Theory

Here a step-by-step process for deriving stock price fragility is provided. We may keep the concepts of fund management in mind, discussed briefly in the introduction; however this theoretical measure may apply to the whole universe of stocks and their respective owners. We begin by defining the change in the total quantity of a share held by a particular fund, holding fixed its price, as the sum of active rebalancing by the fund manager and the remaining flows into and out of the fund (flow-driven trading) respectively. Mathematically, this may be expressed as:

$$P_{it}\Delta n_{ikt} = n_{ikt}P_{it}\left(\frac{\Delta w_{ikt}}{w_{ikt}} - \left(\frac{\Delta P_{it}}{P_{it}} - \sum_j \left(w_{jkt}\frac{\Delta P_{jt}}{P_{jt}}\right)\right)\right) + w_{ikt}f_{kt} \tag{2.1}$$

where n_{ikt} is the number of shares in security i held by fund k at time t, $\Delta n_{ikt} = n_{ikt+1} - n_{ikt}$, P_{it} is the price of security i at time t, f_{kt} represents the net fund inflows for fund k at time t and w_{ikt} is the weight of security i in fund k at time t i.e. the ownership structure of asset i, which may be expressed mathematically as:

$$w_{ikt} = \frac{n_{ikt}P_{it}}{a_{kt}} \tag{2.2}$$

where a_{kt} is the total fund size of fund k at time t. We are only interested in the 2nd term in (2.1) above to construct fragility because the sum of the weighted flows across all funds is assumed to have price impact, as implied earlier. This assumption may be formalized by the following regression equation:

$$r_{it+1} = \alpha + \lambda\frac{\sum_k w_{ikt}f_{kt}}{\theta_{it}} + \varepsilon_{it+1} \tag{2.3}$$

Or in vector notation:

$$r_{it+1} = \alpha + \lambda\frac{W'_{it}F_t}{\theta_{it}} + \varepsilon_{it+1} \tag{2.4}$$

As suggested by Greenwood and Thesmar (2011), the weighted flows are normalized by the stock's market capitalization, θ_{it}. α is the constant term and λ is the regression coefficient representing non-fundamental price impact. The dependent variable r_{it+1} represents asset i's return at time t+1. Returns are expressed as one-step ahead implying a forecasting regression. The error term ε_{it+1} reflects other sources of returns, including news on fundamentals. Greenwood and Thesmar (2011) provide support to the notion that the error term and the regressor are uncorrelated i.e. that weighted flows do not predict future fundamentals. This research project looks at this notion in more detail, discussed in further detail below. Now, taking the variance of (2.4) produces the following regression equation:

$$\sigma^2_{it+1} = \lambda^2\frac{1}{\theta^2_{it}}W'_{it}\Omega_t W_{it} + \sigma^2_{(\varepsilon)it+1} + cov(W'_{it}F_t, \varepsilon_{it+1}) \tag{2.5}$$

where σ^2_{it+1} is the variance of returns, $\sigma^2_{(\varepsilon)it+1}$ is the error variance and Ω_t is the variance-covariance matrix of flows. By Greenwood and Thesmar (2011), the covariance term above is zero. In this research project, the validity of this claim is assessed. The first term in (2.5) captures the fragility measure:

$$G_{it} = \frac{1}{\theta^2_{it}} W'_{it} \Omega_t W_{it} \tag{2.6}$$

Stock price fragility defined above is expressed as a function of an asset's ownership structure as well as the ex-ante variance-covariance matrix of fund flows, normalized by the asset's market capitalization. It may be decomposed into the on-diagonal and off-diagonal terms of the ex-ante variance-covariance matrix as follows:

$$G_{it} = \frac{1}{\theta^2_{it}} W'_{it}(\Omega_t - D_t)W_{it} + \frac{1}{d^2_t} W'_{it}(D_t - d_t I)W_{it} + d_t(mf)^2_{it} H_{it} \tag{2.7}$$

where D_t is the matrix of the diagonal elements of Ω_t, d_t is the mean of these diagonal elements, $(mf)_{it}$ is the proportion of security i held by all funds (the asset's fund ownership) and H_{it} is equivalent to the Hirschman-Herfindahl index (1964) in the sense that it is a measure of ownership concentration of the asset with respect to all funds. The first term in (2.7) represents the off-diagonal elements of Ω_t. Thus, if no correlations exist between fund flows, then this term will be zero. The last two terms represent the on-diagonal elements which are a function of several variables such as fund ownership and ownership concentration.

2.2.2 Application

Having monthly fund sizes and fund returns, we may compute the fund flows as follows:

$$f_{kt} = a_{kt} - a_{kt-1}(1 + R_{kt}) \tag{2.8}$$

where R_{kt} is the return on fund k at time t. From (2.8), it is easy to see that fund flows are proportional to fund sizes. Thus, when computing the variance-covariance matrix of these flows directly, a heteroskedasticity issue arises. Greenwood and Thesmar (2011) suggest estimating this matrix through a rolling window. To do this, the flows are first normalized by the prior month fund size i.e. f_{kt}/a_{kt-1}. Then, to compute Ω_t, the rolling window is implemented from the first time point (August, 2002) up to time t. Finally, Ω_t is rescaled, where $(Da)_t$ is the k-by-k diagonal matrix of fund sizes, as follows:

$$\hat{\Omega}_t = (Da)_t \Omega_t (Da)_t \tag{2.9}$$

In this research project, only a subset of stock owners are dealt with. Thus, an empirical measure of fragility is computed based only on the available data. As mentioned above, this may be a cause for concern in the sense that there may be omitted variable bias. Thus, in this research project, a cautious assessment is followed. Initially, this measure is assessed over time and over all available stocks. To do this, each of its components, such as fund ownership, ownership concentration, flow volatilities and correlations, is separately assessed and compared. Furthermore, an initial inspection of total return volatility and empirical fragility is performed.

After initial inspections, a regression of total returns on weighted flows, as in (2.4), is run in order to look at the correlation between the weighted flows and the error terms, as explained by the third term in (2.5). Then, to validate empirical fragility as a proxy for non-fundamental risk, a regression is run with weighted flows as the dependent variable on empirical fragility. This is followed by several regressions focussed on forecasting the one-step ahead total return volatility by empirical fragility. For a better understanding of this relationship, the decomposed elements of empirical fragility, as in (2.7), are included. In an attempt to control for omitted variable bias and endogeneity, a set of control variables are also included in the regressions suggested by literature. These include: lagged volatility, lagged skewness, firm size, market capitalization, book-to-market ratio and lagged share turnover.

All regressions implement the Fama-Macbeth first-stage forecasting regression, save one, where a panel regression is run with firm fixed effects in an attempt to further control for endogeneity (please refer to the literature review). In each Fama-Macbeth regression Newey-West robust standard errors are provided (1987), explained earlier. In this project, we do not explicitly discuss the extent of the autocorrelation and heteroskedasticity present in the regressions. Thus, we make the intuitive assumption that a lag length of 2 is most appropriate as seen in the literature. The covariance matrix for the standard errors may be expressed mathematically as:

$$\sigma_e^2(X'X)^{-1} = \frac{n}{n-k}\sum e_t^2 x_t' x_t + \frac{n}{n-k}\sum_{l=1}^{p}\left(1-\frac{l}{p+1}\right)\sum_{t=l+1}^{n} e_t e_{t-1}(x_t' x_{t-l} + x_{t-l}' x_t) \quad (2.10)$$

Where σ_e^2 is the variance of the error terms, l represents the lag length, e_t is the error term at time t and x_t the observations at time t.

All observations are equally weighted in each cross-section, save one, where observations are weighted by their fund ownership to control for measurement error in those stocks with low fund ownership. Finally, the Fama-French one- three- and four- factor models (Fama and French, 1992; 1993; Carhart, 1997) are implemented with excess return volatility on empirical fragility. Mathematically, where Z_{it} represents the matrix of controls, all regressions may be formalized as follows:

$$\sigma_{it+1} = a + b\sqrt{G_{it}} + Z_{it}C + u_{it+1} \tag{2.11}$$

2.3 Co-fragility

2.3.1 Theory

Co-fragility may be defined as the co-movement of the flows of two assets. Formalizing this measure requires taking the covariance of (2.4) to produce:

$$cov_t(r_{it+1}, r_{jt+1}) = \frac{\lambda^2}{\theta_{it}\theta_{jt}} W'_{it}\Omega_t W_{jt} + cov(\varepsilon_{it+1}, \varepsilon_{jt+1}) \tag{2.12}$$

As in Greenwood and Thesmar (2011), the assumption here is that the covariance between the weighted flows and the error terms are zero. Thus, (2.12) has two terms, the first of which is the mathematical expression for co-fragility. To reiterate:

$$G_{ijt} = \frac{W'_{it}\Omega_t W_{jt}}{\theta_{it}\theta_{jt}} \tag{2.13}$$

Having mastered the theory behind stock price fragility, co-fragility, (2.13), is self-explanatory and should predict the covariances of total returns. We may also go one step further by predicting the correlations of total returns. To do this, co-fragility is normalized to produce:

$$G^{norm}_{ijt} = \frac{G_{ijt}}{\sqrt{G_{it}G_{jt}}} \tag{2.14}$$

2.3.2 Applications

As was done in the previous section, an empirical measure of co-fragility is constructed based on the available data. An initial inspection is also performed by graphically assessing the relationship between co-fragility and the covariances of total returns as well as normalized co-fragility with the correlations of total returns. This is followed by several regressions focussing on this relationship with the same techniques used as in the previous section. The controls include industry-specific factors, firms with similar size, book-to-market ratios and number of similar owners (logged), and lagged covariances and correlations of total returns. For industry-specific factors, dummy variables are constructed whereby one is given to stocks with the same SIC codes and zero otherwise. There are four dummy variables in this case representing the one-, two-, three- and four-digit SIC codes. The "similarity" variables, for book-to-market ratios, number of owners and firm size, are constructed by taking and storing the difference between the values of each of these for each pair of stocks.

Mathematically all regressions for predicting covariances and correlations of total returns respectively may be formalized as follows:

$$\sigma_{ijt+1} = a + bG_{ijt} + Z_{ijt}C + u_{ijt+1} \tag{2.15}$$

$$\rho_{ijt+1} = a + bG_{ijt}^{norm} + Z_{ijt}C + u_{ijt+1} \tag{2.16}$$

2.4 Fragility Beta

2.4.1 Theory

The fragility beta may be defined as the sensitivity of the flows into an asset relative to the flows into a portfolio of assets. Mathematically, the fragility beta of an asset i at time t with respect to portfolio p, where s_{it}^{p} represents the weight of an asset i at time t in the portfolio, is:

$$G_{it}^{p} = \frac{\sum_j \left[s_{jt}^{p} G_{ijt}\right]}{\left[\sum_{j,j'} s_{jt}^{p} s_{j't}^{p} G_{jj't}\right]} \tag{2.17}$$

2.4.2 Applications

In this research project, the relationship between fragility betas and total return betas is assessed relative to three benchmark portfolios: an equally weighted portfolio, a HML-weighted portfolio (i.e. the stock with the highest book-to-market ratio is assigned the largest weight in the portfolio) and a SMB-weighted portfolio (i.e. the stock with the biggest market capitalization is assigned the largest weight in the portfolio). An important point to note is that although it may appear counter-intuitive to apply the SMB-weighted portfolio - considering that our data set contains only 41 of the largest stocks - it is hypothesized that the difference between the market capitalisations of each stock in this set is sufficient to obtain a size effect. Specifically, we note our biggest stocks (e.g. SAB and BIL) are roughly ten times bigger than our smallest stocks (e.g. MPC and ASR).

This assessment is performed graphically as well as with the use of several regressions based on the techniques discussed previously. The controls used include fund ownership, number of owners, book-to-market ratios, firm size as well as market capitalization. Mathematically, all regressions may be formalized as follows:

$$\beta^p_{it+1} = a + bG^p_{it} + Z_{it}C + u_{it+1} \tag{2.18}$$

2.5 Fragility & Arbitrage

2.5.1 Theory

As stated in the literature review, this research project attempts to assess whether arbitrageurs are able to accommodate the liquidity shocks of funds through the order imbalances driven by the active trades of the funds. To do this, equation (2.4) must be adjusted to allow for the additional impact of active trades on total returns. This may be mathematically formalized as follows:

$$r_{it+1} = \alpha + \lambda \frac{W'_{it}F_t}{\theta_{it}} + \lambda D^X_{it} + \varepsilon_{it+1} \tag{2.19}$$

where D^X_{it} represents the order imbalances driven by active trades and may also be expressed as follows:

$$D^X_{it} = \delta_{it} + \gamma_{it} \frac{W'_{it}F_t}{\theta_{it}} + \upsilon_{it} \tag{2.20}$$

From (2.20), it can be deduced that if $\gamma_{it} < 0$ then arbitrageurs are able to mitigate the effect that flow-induced trading has on total returns. If $\gamma_{it} > 0$ then arbitrageurs add to the price impact of flow-induced trading. Now, substituting (2.20) into (2.19) and taking the variance (assuming independence between G_{it} and ε_{it+1}) leads to:

$$\sigma^2_{it+1} = \lambda^2(1+\gamma_{it})^2 G^2_{it} + var_t\Big(\varepsilon_{it+1} + \lambda \upsilon_{it}\Big) \tag{2.21}$$

From (2.21), it may be deduced that the coefficient of $(1+\gamma_{it})^2$ or $|1+\gamma_{it}|$ is a measure of the sensitivity of total return volatility to stock price fragility. Thus, if either of these terms are small, then it is expected that stock price fragility has little or no impact on return volatility, and vice versa.

2.5.2 Applications

With reference to the aforementioned theory, testing for the impact of arbitrageurs on flow-induced trading may be performed in two stages. First, a regression is run based on (2.20) above and the estimates of γ_{it} are assessed and stored. Second, the following regression is run and assessed:

$$\sigma^2_{it+1} = a + b|1 + \gamma_{it}| + c\sqrt{G_{it}} + d|1 + \gamma_{it}| \cdot \sqrt{G_{it}} + u_{it+1} \tag{2.22}$$

As has been followed throughout the methodology, the equally-weighted Fama-Macbeth regression technique, with and without the Newey-West estimator, is used.

3

Results

This chapter may be broken down into four sections. The first section is concerned with the results associated with stock price fragility. Specifically, we look at fragility and its components through graphical and tabular inspections as well as several regressions estimates predicting total volatility. The second and third sections explain the results associated with co-fragility and fragility beta respectively, in a similar light to the first section. Finally, the results pertaining to the relationship between fragility and arbitrage is discussed.

3.1 Stock Price Fragility

Following the methodology discussed above, fragility and its components are presented graphically below. Figure 3.1 illustrates that the number of fund owners for the median share per time point remains relatively stable over time but decreases slightly beginning during the period of the 2007 financial crisis till present. We would normally expect the number of owners to increase over time however evidence seems to point otherwise. A potential explanation for this observation is saturated ownership at the beginning of the 10 year period under consideration, with a decrease in ownership as foreigners exited emerging markets during the ensuing risk-off period. We support this notion with figure 3.2 which illustrates the proportion of the median share per time point held by all funds relative to all shares of that stock in the market (mf_{it}); despite the sharp drop during 2007 which may indicate that the global financial crisis had an impact on the ownership structure of funds across all stocks. The latter notion approximately corresponds to figure 3.3 whereby the concentration of fund ownership (H_{it}) for the median share per time point increases beginning during the same period till present. It must also be noted that in a developing market, such as in SA, we expect a higher concentration of ownership. Finally, figure 3.4 indicates that fragility has risen beginning during the period of the 2007 financial crisis to present. Keeping the breakdown of stock price fragility in mind, it seems that the effect that fund ownership has on fragility is offset by ownership concentration. Thus, the flow volatilities and correlations may be the culprit. We may also intuitively suggest that flow volatility is larger and that the correlations of flows is more apparent over periods of crisis. Although these results do not correspond to those of Greenwood and Thesmar (2011), we note that the results found in a developed market context may not directly translate to an emerging market, due to the observed liquidity differences and higher ownership concentrations.

Figure 3.1: The number of fund owners for the median share per time point

Figure 3.2: The proportion of the median share per time point held by all funds

Figure 3.3: The ownership concentration for the median share per time point

Figure 3.4: Stock price fragility for the median share per time point

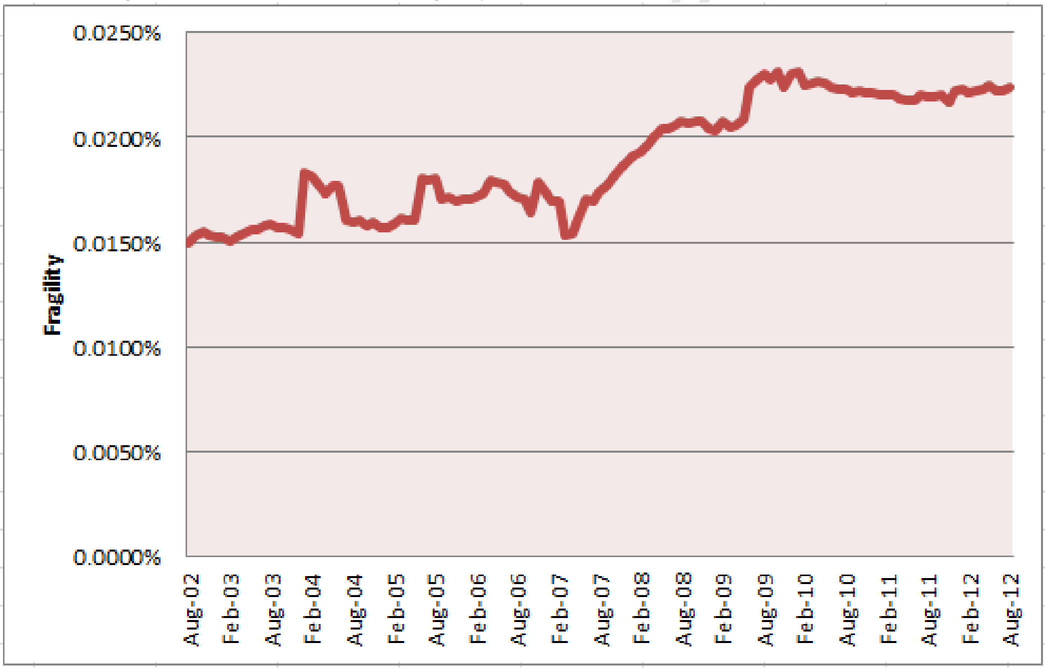


Table 3.1 below explores the fragility and component estimates further. The first three components illustrate the ownership structures of the funds for the mean share as well as the shares per quartile. Given the fact that we have sufficient data for the ownership structures of the funds, there seems to be a relatively stable quartile spread for ownership concentration, fund ownership and number of owners. Reinforcing the graphical depiction of these components,

the lower quartiles resemble the graphs when comparing each time point. However, at higher quartiles it seems that there are several outliers pulling away from the common patterns. We take note of this before regressing each individual component against total volatility. The last two components are concerned with the flows of the funds per share mean and quartile. Considering that we had insufficient data, it is understandable that at lower quartiles the values for the volatilities and correlation flows are insignificant. However, as we move up to higher quartiles, the results seem correspond to the graphs above i.e. from the crisis period onwards, there is an upward shift in fragility estimates.

Fragility & Components	**Date**	**Mean**	**Min**	**25%**	**Median**	**75%**	**Max**
Ownership Concentration (H_{it})	Aug-2012	0.15	0.04	0.08	0.11	0.17	0.69
	Aug-2007	0.16	0.03	0.05	0.07	0.20	0.76
	Aug-2002	0.14	0.04	0.06	0.09	0.16	0.61
Fund Ownership ($(mf)_t$)	Aug-2012	0.46	0.09	0.21	0.43	0.72	0.93
	Aug-2007	0.4	0.05	0.22	0.33	0.68	0.86
	Aug-2002	0.43	0.08	0.32	0.39	0.6	0.83
Number of Owners	Aug-2012	133.59	27	93	126	178	242
	Aug-2007	120.46	24	69	130	161	235
	Aug-2002	127.76	8	89	130	168	285
Flow Volatility (σ^f_{kt})	Aug-2012	0.09	0.00	0.01	0.10	0.27	0.48
	Aug-2007	0.10	0.00	0.02	0.11	0.28	0.54
	Aug-2002	0.07	0.00	0.00	0.08	0.22	0.44
Flow Correlation ($\rho^f_{kk't}$)	Aug-2012	0.06	0.00	0.00	0.04	0.07	0.11
	Aug-2007	0.06	0.00	0.00	0.05	0.07	0.10
	Aug-2002	0.04	0.00	0.00	0.03	0.05	0.08
Fragility (G_{it}) $\times 10^{-3}$	Aug-2012	0.09	0.00	0.00	0.20	0.27	0.41
	Aug-2007	0.09	0.00	0.00	0.17	0.29	0.48
	Aug-2002	0.07	0.00	0.00	0.15	0.22	0.34

Table 3.1: Table showing stock price fragility and its components for the mean share as well the shares per quartile at 3 different time points

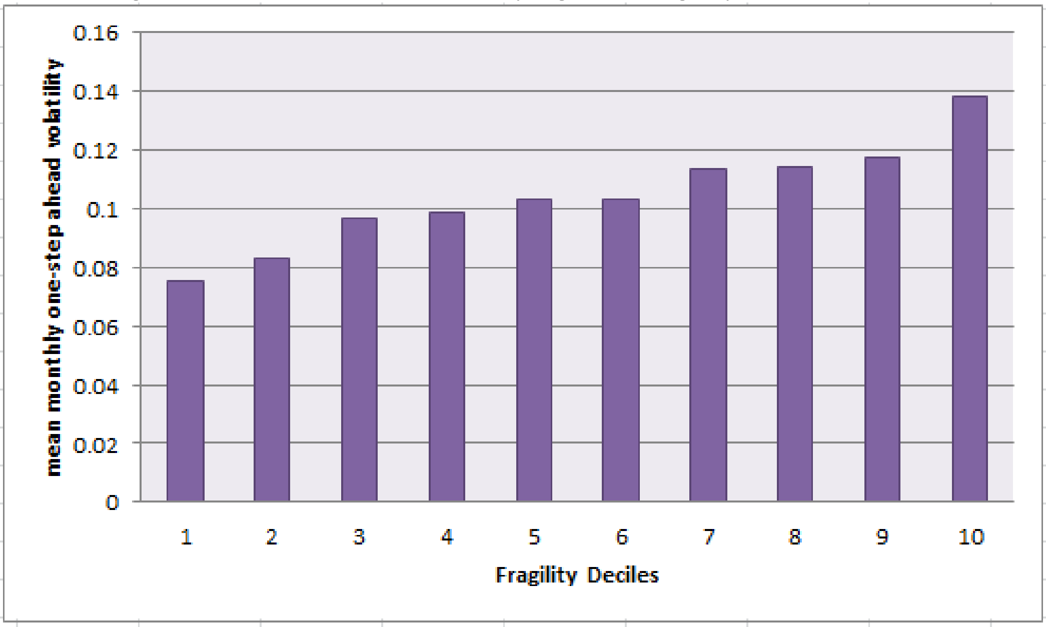
Table 3.2 below illustrates those variables correlated with fragility at three different time points. These correlations are split into quartiles by the size of the correlation of each share in order to understand how the components of fragility influence the measure across stocks. What is discernible from this table is that fragility is not constant across stocks. More specifically, the drivers of fragility differ across stocks and time points. This result is intuitive in the sense that there is no single driver of fragility i.e. each component could have a notable impact on fragility depending on the stock. Another interesting point that must be made is that some form correlation does exist between total volatility and fragility. The only issue we may raise here is the fact that the average correlations of each component are close to zero, save flow volatility. This result is unusual considering that our data issues are, for the most part, attributed to flow data. This may suggest that our major concern is not the bias in flow volatility alone but also in the fragility measure as a whole.

Fragility Correlates	**Date**	**Mean**	**Min**	**25%**	**Median**	**75%**	**Max**
Ownership Concentration (H_{it})	Aug-2012	0.013	-0.131	-0.104	0.108	0.127	0.190
	Aug-2007	0.006	-0.154	-0.107	0.110	0.122	0.172
	Aug-2002	0.006	-0.145	-0.109	0.104	0.124	0.164
Fund Ownership ($(mf)_t$)	Aug-2012	0.019	-0.119	0.103	0.120	0.135	0.159
	Aug-2007	0.018	-0.132	0.102	0.115	0.142	0.074
	Aug-2002	0.007	-0.152	-0.110	0.102	0.117	0.167
Number of Owners	Aug-2012	0.013	-0.115	-0.101	0.116	0.125	0.141
	Aug-2007	0.015	-0.117	-0.101	0.118	0.131	0.146
	Aug-2002	0.014	-0.116	0.000	0.118	0.132	0.143
Flow Volatility (σ^f_{kt})	Aug-2012	0.176	0.093	0.169	0.179	0.190	0.208
	Aug-2007	0.181	0.147	0.172	0.183	0.194	0.213
	Aug-2002	0.177	0.133	0.167	0.179	0.187	0.228
Flow Correlation ($\rho^f_{kk't}$)	Aug-2012	-0.002	-0.042	-0.015	-0.004	0.012	0.062
	Aug-2007	-0.007	-0.048	-0.024	-0.008	0.013	0.051
	Aug-2002	0.000	-0.056	-0.016	-0.001	0.013	0.056
Lagged Fragility (G_{it-1})	Aug-2012	0.012	-0.145	-0.120	-0.123	0.119	0.166
	Aug-2007	0.043	-0.167	-0.115	-0.001	0.024	0.258
	Aug-2002	0.002	-0.129	-0.114	0.102	0.116	0.142
Return Volatility (σ^r_{kt})	Aug-2012	0.006	-0.148	-0.117	0.106	0.125	0.162
	Aug-2007	-0.091	-0.185	-0.115	-0.104	0.121	0.148
	Aug-2002	0.002	-0.169	-0.128	0.101	0.118	0.185

Table 3.2: Table showing the correlation of a number of variables with stock price fragility for shares on each quartile at three different time points

From figure 3.5 below, we attempt to acquire a basic idea of the relationship between fragility and total volatility. To construct the graph, fragility is sorted into deciles and the average of each decile is taken. It seems from the graph that total volatility is approximately directly proportional to fragility; a result we would expect. However, we cannot infer here that this relationship exists i.e. given our small sample size, this result may be purely coincidental. Moreover, we only have 41 stocks available in our sample so that we have approximately 4 stocks per decile. Thus, further investigation into the relationship between total volatility and fragility is required - we do so by running several regressions, the results of which are depicted in the table 3.6 below.

Figure 3.5: Total return volatility against fragility sorted in deciles



Keeping the standard methodology behind the regressions in mind (see the previous chapter), we note the following from table 3.6. Firstly, the predictors are in each row alongside their associated beta coefficients (in black) and t-statistics below that (in red). Each column represents a different regression. Additional information pertaining to the regressions are provided at the bottom of the table. In column 1 ("Regression 1") we perform a Fama-Macbeth regression, where observations at each cross section are equally weighted, with one-step ahead total return volatility is regressed against those components of fragility regarding the ownership structure of the funds (because we have adequate ownership structure data) as well as the major controls of total volatility suggested by literature which include: lagged return volatility and skewness, the scaled log of book value and market capitalization, and the book-to-market ratio. The results from Regression 1 indicate that all coefficients are significant, with a high r-squared, save the scaled log of book value and market capitalization, which may be due to multicollinearity present in the model i.e. book value, market value and book-to-market ratio are likely to be highly correlated with one another. Nevertheless, the significance of book-to-market ratio supports existing literature in both the international and SA market. The significance of the remaining controls also support existing literature. What is most interesting about Regression 1 is that the ownership structure of SA funds appears to have a small impact on total SA return volatility. Intuitively, we would expect the total ownership structure of a stock market to be a factor in determining total return volatility, but because only a subset of ownership is available (select funds) this notion should be difficult to prove. However, it seems our subset is sufficient to proxy the total SA market.

Figure 3.6: Several regressions of total return volatility against fragility and other controls

	Total Return Volatility: $\sigma_{i,t+1}$							Excess Return Volatility: $\sigma_{i,t+1}$		
Predictors/Regressions	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant	0.01081	0.13499	0.13534	0.12311	0.01230	0.01002	0.01230	0.01203	0.01256	0.01567
	13.76483	15.34432	18.31123	15.32349	12.95030	8.28893	12.93389	16.28945	17.26783	17.27846
$\sqrt{G_{i,t}}$		1.43244	1.63425		0.72250	0.77489	0.72250	1.12678	1.10384	0.98838
		10.90430	11.23645		2.43799	1.13242	2.72682	1.26834	1.24843	1.17839
$\sqrt{G_{i,t}}$ (on-diag)				0.55428						
				1.26752						
$\sqrt{G_{i,t}}$ (off-diag)				1.03546						
				16.67839						
Log(# owners)$_t$	0.00258				0.00264	0.00784	0.00264			
	14.28793				11.26372	9.47848	10.48821			
(mf)$_t$	-0.00207				-0.00206	-0.00348	-0.00206			
	-9.36734				-7.75894	-5.48330	-6.39024			
$H_{i,t}$	-0.01116				-0.01215	-0.01329	-0.01215			
	-11.23564				-7.67832	-6.48839	-6.94924			
Lagged Volatility	0.78442				0.77911	0.74930	0.77911			
	36.72891				27.38930	24.37832	26.39942			
Lagged Skewness	-0.00005				-0.00005	-0.00005	-0.00005			
	-21.34632				-18.89302	-18.78920	-17.37832			
log(Firm Size)/100	0.04727				0.04845	0.04789	0.04845			
	1.28931				1.23859	1.18299	1.17829			
log(Market Cap)/100	-0.04913				-0.04766	-0.04788	-0.04766			
	-1.45635				-1.23785	-1.24895	-1.14323			
Book/Market	0.00303				0.00304	0.00302	0.00304			
	34.72637				33.28894	32.38992	32.67382			
R^2	0.64	0.12	0.15	0.11	0.67	0.65	0.67	0.05	0.06	0.08
Regression Technique	FM	FM	FM	FM	FM	Panel FE	FM (NW)	FM	FM	FM
Weighting	EW	EW	MF	EW	EW	EW	EW	EW	EW	EW

Regression 2 performs the same type of regression yet with only one predictor, the square root of our estimated fragility measure. The associated coefficient is significant, further supporting the relationship between these variables depicted in figure 3.5. Yet, as previously mentioned, without a thorough investigation, this relationship may be purely coincidental.

Regression 3 retains the same variables as Regression 2 but, here, we weight each stock by their mutual fund share. This regression is performed to simply mitigate the measurement error in the model i.e. down-weighting those shares taking up only a small portion of the funds. This way, empirical fragility serves as a better proxy to true fragility. Although the coefficient on fragility is made more significant by this adjustment, it is not a notable improvement. Moreover, the data issues pertaining to our fragility measure is enough to suggest that this result is negligible.

Regression 4 follows the same idea as Regression 2 except here, fragility is broken down into it's on- and off-diagonal components. This regression allows us to understand the cause of the relationship between total return volatility and fragility based on Regression 2. It is clear that the off-diagonal elements are the drivers. This may be due to the data issues pertaining to flows alone. Thus, the ownership structure of the funds is the driver. We may also suggest that during the financial crisis period and beyond, the correlations of flows were a lot higher resulting in common movements between return volatility and fragility.

Regression 5 is an equally-weighted Fama-Macbeth regression involving all available predictors. The significance of the control coefficients remain significant and, relative to Regression 1, these results are to be expected. With regards to the fragility coefficient, it has significantly weakened relative to Regression 2. This is due to the fact that the ownership structure of the funds are the primary drivers of total return volatility as seen by the significance in the ownership concentration, number of owners and mutual fund share coefficients. This result is also due to the relationship between fundamental and non-fundamental demand i.e. by controlling for fundamentals, fragility is weakened and therefore we may have eliminated some endogeneity within the model. However, we point out that we cannot measure the differences in ownership endogeneity here; we only consider the possibility of it existing. Nevertheless, the beta coefficient regarding fragility approximately corresponds to that of Greenwood and Thesmar (2011).

Regression 6 further investigates ownership endogeneity by performing a panel regression with firm fixed effects (see the literature review for more details). The significance of the coefficients are weakened by this regression and therefore we cannot suggest that endogeneity has been sufficiently controlled for here. We may only suggest that random effects is best suited for this type of model.

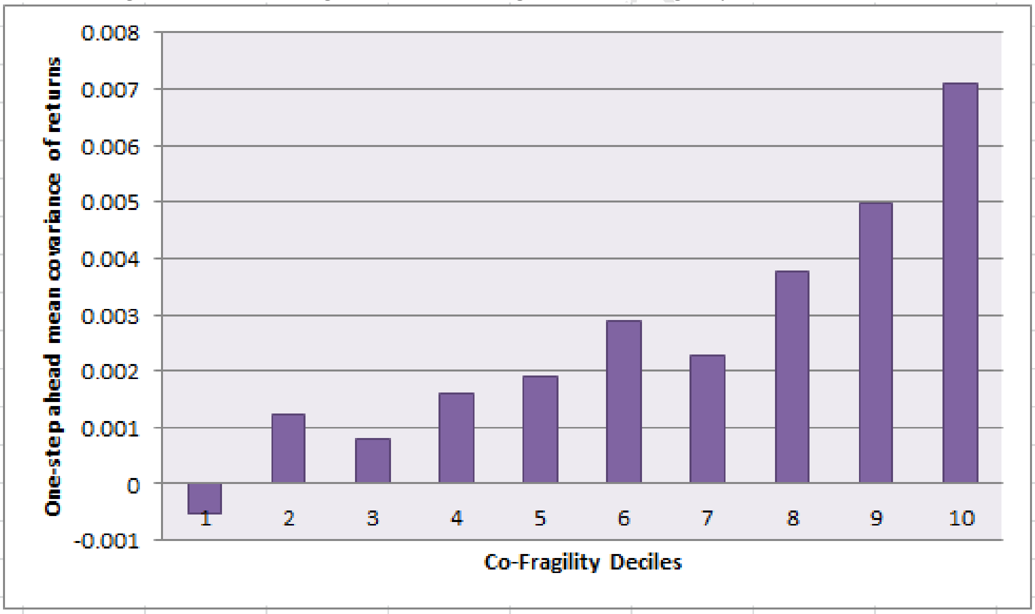
Regression 7 is the same as Regression 5, save the additional Newey-West estimator (1987) to adjust the error terms, and ultimately the standard errors of the coefficients, for potential autocorrelation and heteroskedasticity in the model (see the literature review for more details). Theoretically, it is expected that the significance of the beta coefficients are weakened by the introduction of the Newey-West estimator. Although we do not show in this project whether there exists heteroskedasticity or autocorrelation in the model, ideally we would want the coefficients to remain significant with robust standard errors. In this regression this seems to be the case for all those coefficients that were significant in Regression 5. The unusual piece is the increase in significance of the fragility coefficient. This could be due to a negative correlation in the error terms. It could also be due to the fact that the model has been misspecified with the imposition of the fragility measure in the model. This would make sense considering the data issues we have faced. For those that weren't originally significant, there is no notable improvement.

Regressions 8,9 and 10 represent the Fama-French one- three- and four- factor models (Fama and French, 1992; 1993; Carhart, 1997) respectively, whereby excess return volatility is regressed on empirical fragility. Compared to the previous regressions, the significance of the beta coefficients on fragility have weakened. We posit empirical fragility to be related to total volatility rather than excess volatility. Thus, this result was expected and we may rule out the latter relationship.

3.2 Co-Fragility

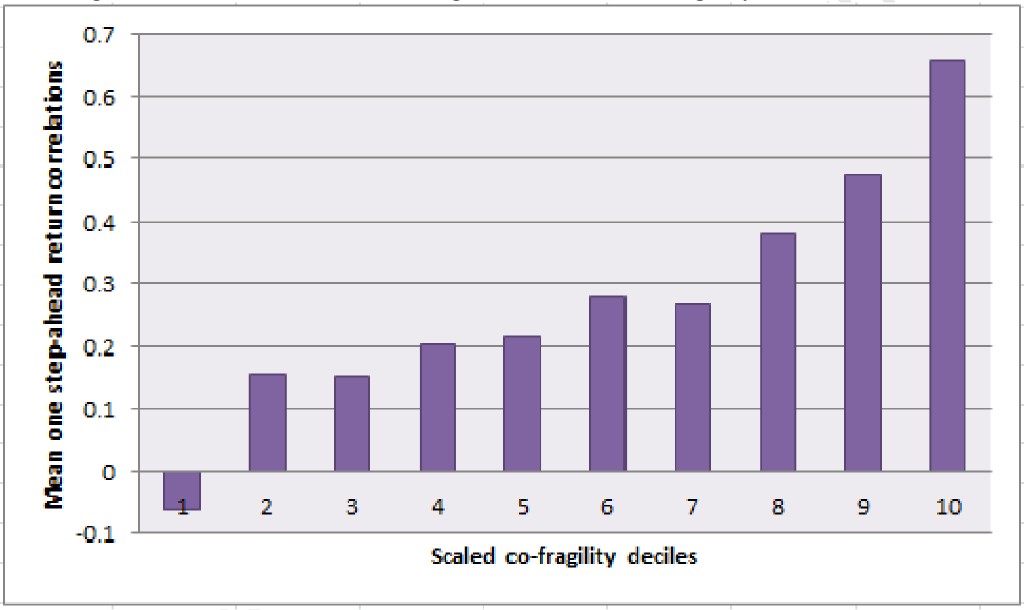
The focus of this section is a discussion of the results pertaining to the relationship between empirical co-fragility and total return covariance. We initially perform a graphical inspection of this relationship. This is followed by a more thorough assessment based on several regression techniques.

Figure 3.7: Average covariance against co-fragility sorted in deciles



From figure 3.7 above, we attempt to acquire a basic idea of the relationship between co-fragility and the average covariance of returns. As we did in figure 3.5 above, we construct the graph by sorting co-fragility into deciles and the average in each decile is then plotted against the average covariance of returns. As we saw above, it appears that average covariance is proportional to co-fragility; a result we would expect. Again however, we cannot infer here that this relationship exists i.e. given our small sample size, this result may be purely coincidental. In order to further investigate this relationship, we run several regressions, see later. The same argument applies to figure 3.8. It is interesting to note in the latter figure that correlations are, for the most part, high and positive. Although we would expect the correlation of returns to be more positive, it seems unusual that return correlations are high. Moreover, one would intuitively expect return correlations to be lower in an emerging market, such as South Africa, than in a fully developed one, like the US. However, comparatively speaking, this result does not correspond to that of Greenwood and Thesmar (2011). There could be several reasons for this result. Firstly, our data contains a small number of the largest capped stocks - reinforcing the notion of bias in our data set. Secondly, the crisis period may have caused stock prices to move in unison. Although we have digressed slightly, this information may be of use in the regressions, see below.

Figure 3.8: Total correlation against scaled co-fragility sorted in deciles



In table 3.9 we provide the results concerning the one-step ahead prediction of the covariance of returns (columns 1 through to 4) and the correlation of returns (columns 5 through to 8). In column 1 ("Regression 1") we perform a Fama-Macbeth regression, where observations at each cross section are equally weighted, and the covariance of returns are regressed only against controls including industry dummy variables, lagged covariance of returns as well as dummy variables for stocks with similar firm sizes, similar number of owners (funds - logged) in all funds and similar book-to-market ratios. The results from Regression 1 indicate that those stocks forming part of the same major division and division industries significantly contribute to the covariance of returns whereas as those in major group and group industries do not. This result is expected - it can be attributed to the fact that proxy funds were constructed by industry (as explained in the data section earlier) and as we move to a smaller group, we deal more with firm specific characteristics whose risks are more likely to be diversifiable. This notion is supported by the fact that we have only a subset of 41 stocks i.e. it may be the case that the group industries in this sample contain only 2 stocks. The other controls remain significant, supporting existing international and South African literature, save the scaled log of book value, which may again be due to multicollinearity present in the model i.e. book value and book-to-market ratio are likely to be highly correlated with one another. Most importantly here though, the ownership structure of funds significantly contribute to the covariance of returns.

Figure 3.9: Several regressions of the average return covariance against co-fragility, scaled co-fragility and other controls

	Total Return Monthly Covariance: σ_{ijt+1}				Total Monthly Return Correlation: ρ_{ijt+1}			
Predictors/Regressions	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	-0.023664	-0.02458	-0.02759	-0.02759	-0.18399	-0.17892	-0.16289	-0.16289
	-9.38920	-7.38992	-8.38902	-6.11627	-14.28849	-13.22343	-15.27482	-14.21131
G_{ijt}		4.15262	2.13829	2.13829				
		9.38892	1.27783	1.46381				
$G_{ijt}/\sqrt{G_{it}}\sqrt{G_{jt}}$						1.27382	0.32920	0.32920
						8.38201	2.62889	3.09010
$SIC1_{it} = SIC1_{jt}$	0.56483		0.51789	0.51789	0.12671		0.11367	0.11367
	5.37281		5.85021	5.49020	8.49201		8.29021	8.14242
$SIC2_{it} = SIC2_{jt}$	0.48929		0.47859	0.47859	0.11300		0.10303	0.10303
	4.58993		4.95003	4.33002	5.39021		5.41884	5.39022
$SIC3_{it} = SIC3_{jt}$	0.10300		0.12002	0.12002	0.08790		0.09579	0.09579
	0.94783		1.00392	0.99305	1.74993		1.58932	1.43902
$SIC4_{it} = SIC4_{jt}$	0.00300		0.00378	0.00378	0.00849		0.00876	0.00876
	0.18394		0.17499	0.16300	0.27493		0.26488	0.19849
Lagged Covariance	0.43829		0.41890	0.41890				
	7.49003		8.58930	8.14022				
Lagged Correlation					0.38493		0.37882	0.37882
					4.28919		5.38012	5.20102
Similar Firm Size (log/100)	-0.02784		-0.02889	-0.02889	-0.21849		-0.23893	-0.23893
	-0.94003		-0.58930	-0.44903	-0.89503		-0.75893	-0.68493
Similar # Owners log	0.03267		0.03192	0.03192	0.28203		0.21893	0.21893
	2.48399		2.14890	2.04993	2.83930		2.95031	2.92892
Similar Book/Market	0.04773		0.04678	0.04678	0.37199		0.32110	0.32110
	4.29902		3.90503	3.39294	2.89403		2.90204	2.12903
R^2	0.2	0.03	0.22	0.22	0.18	0.04	0.19	0.19
Regression Technique	FM	FM	FM	FM (NW)	FM	FM	FM	FM (NW)
Weighting	EW	EW	EW	EW	EW	EW	EW	EW

Regression 2 performs the same type of regression yet with only one predictor, our co-fragility measure. The associated coefficient is significant, further supporting the relationship between these variables depicted in figure 3.7 above. Yet, as previously mentioned, without a thorough investigation, this relationship may be purely coincidental.

Regression 3 includes all available predictors. The significance of the control coefficients remain similar to that of Regression 1 - these results are to be expected. With regards to the co-fragility coefficient, it has significantly weakened relative to Regression 2. This is partly due to the ownership structure of the funds being the primary drivers of co-fragility as seen by the significance in the log of the number of similar owners. However, this result is mostly due to endogeneity within the model. Greenwood and Thesmar (2011) illustrate the significance of co-fragility in their regressions, yet we do not. We had expected there to be a stronger significance for reasons provided in the literature review above. Apart from the aforementioned reasons to this, we may infer that the insignificance of the beta coefficient on co-fragility may be due to data problems.

Regression 4 is the same as Regression 3, save the additional Newey-West estimator (1987). As expected, the significance of the beta coefficients are weakened by the introduction of the Newey-West estimator. Although we do not show in this project whether there exists heteroskedasticity or autocorrelation in the model, ideally we would want to the coefficients to remain significant with robust standard errors. In this regression this seems to be the case. For those that were not significant in the previous regression, there is no notable improvement in significance.

In regression 5 we begin regressing the correlation of returns against the same controls as Regression 1 except here we replace the lagged covariance of returns with the lagged correlation of returns. It's clear here that the results are very similar to that of Regression 1. The only major difference is the decrease in the overall size of the coefficients, compensated by the increase in the absolute size of the constant coefficient. This, of course does not explain the high correlations explained through figure 3.8 above - there may be a missing coefficient here (omitted variable bias). Though the controls have less impact, it still is significant overall. We may thus apply the same reasoning as we did in Regression 1 above to explain the significance of the coefficients.

Regression 6 relates to Regression 2 except we replace co-fragility with scaled co-fragility. Again, the associated coefficient is significant, further supporting the relationship between these variables depicted in figure 3.8. Yet, as previously mentioned, without a thorough investigation, this relationship may be purely coincidental.

Regression 7 replicates Regression 3 with scaled co-fragility and lagged correlations of returns instead. The results are very similar to Regression 3. The most notable difference is scaled co-fragility remaining significant even after controlling for endogeneity. Considering that fragility is significant and co-fragility is not, it may be the case that fragility dominates in the calculation of scaled co-fragility i.e. a small change in fragility may result in a larger change in scaled co-fragility as opposed to a small change in co-fragility.

Regression 8 implements the Newey-West estimator (1987). As we have noted before, the significance of the beta coefficients are weakened by the introduction of the Newey-West estimator. Those that are significant weaken only slightly, save the increase in significance of the scaled co-fragility coefficient, for reasons explained above. Again, for those that were not significant in the previous regression, there is no notable improvement in significance.

3.3 Fragility Beta

The focus of this section is a discussion of the results pertaining to the relationship between the empirical fragility beta and total return betas based on the three benchmark portfolios discussed previously. We initially perform a graphical inspection of this relationship. This is followed by a more thorough assessment based on several regression techniques.

From figure 3.10, 3.11 and 3.12 below, we attempt to acquire a basic idea of the relationship between the fragility beta and the beta of returns based on market, HML and SMB benchmark portfolios respectively. As we did in figure 3.5 above, we construct the graph by sorting the fragility betas into deciles and the average in each decile is then plotted against the average betas of returns, weighted by the respective benchmark portfolios. Unlike we saw above, it appears that there is little or no relationship between the fragility betas and the betas of returns. Data complications aside, this result is unusual considering that we've shown in the above figures that there is a positive relationship between fragility and the volatilities of returns and between co-fragility and the covariances of returns. One possible reason for this is that the betas of returns encompass 100% weight in the 41 selected stocks whereas the fragility betas encompass that portion of the selected stocks owned by funds. Another possible reason is that the benchmark portfolio for the beta of returns tends to adequately reflect fundamental market movements whereas the fragility benchmark is based on the weighted flows into and out of funds. Thus, from these figures we expect there to be little or no significance in the regression coefficients of the fragility betas.

Figure 3.10: Market betas against market fragility betas sorted in deciles

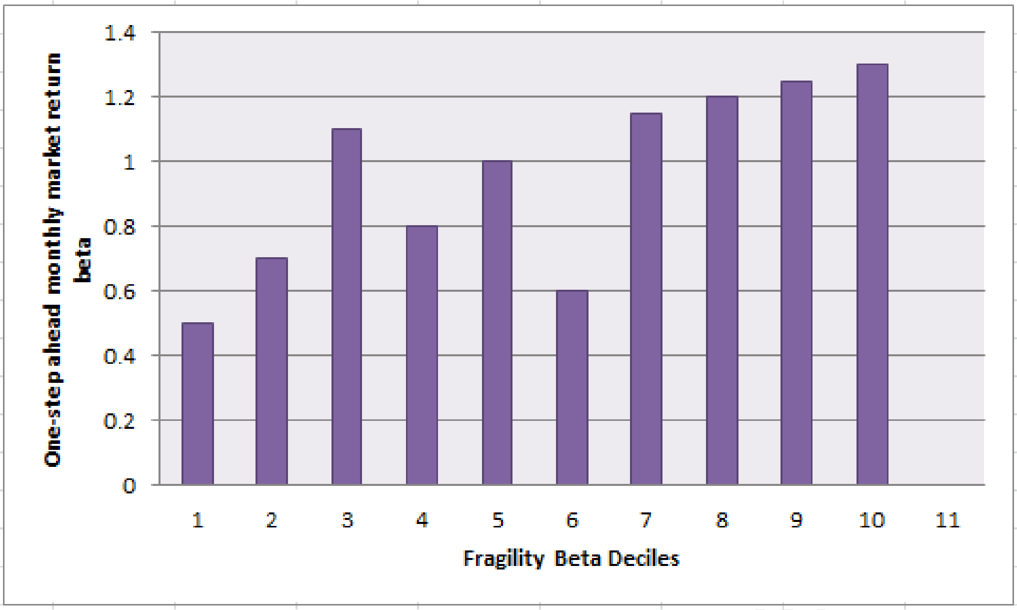


Figure 3.11: HML betas against HML fragility betas sorted in deciles

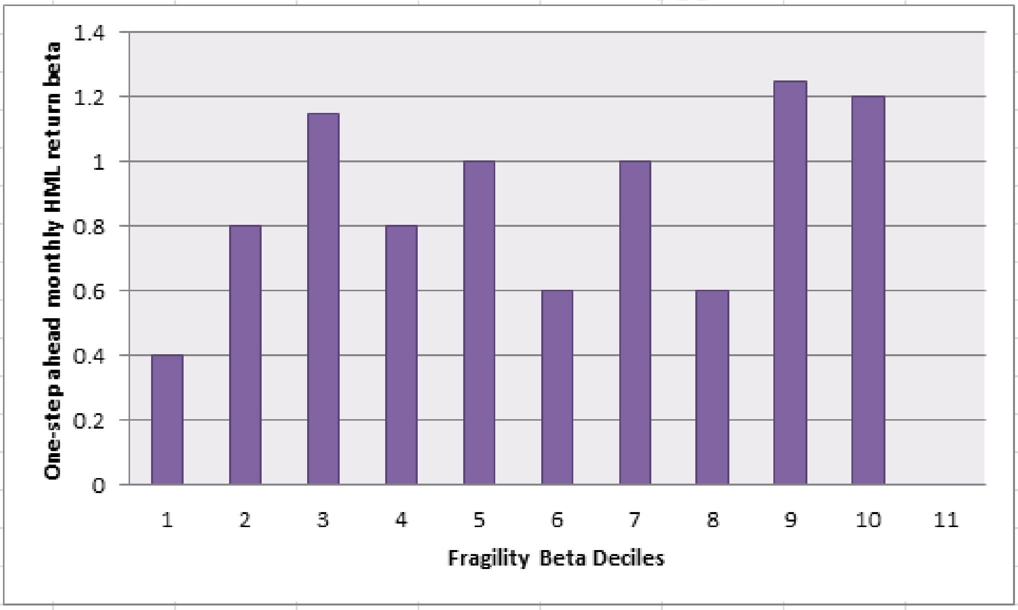
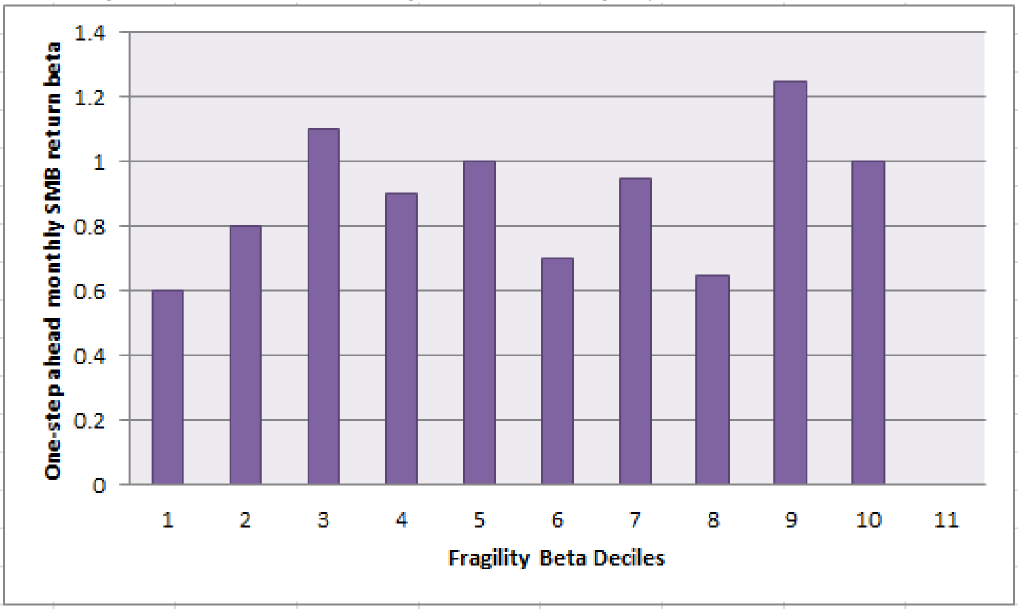


Figure 3.12: SMB betas against SMB fragility betas sorted in deciles



In table 3.13 we provide the results concerning the one-step ahead prediction of the market betas of returns (columns 1 through to 3), the HML betas of returns (columns 4 through to 6) and the SMB betas of returns (columns 7 through to 9). In column 1 ("Regression 1") we perform a Fama-Macbeth regression, where observations at each cross section are equally weighted, and the market betas are regressed only against controls including the log of the number owners (funds), the mutual fund shares of stocks, the scaled log of book value and market capitalization, and the book-to-market ratio. The results from Regression 1 indicate that there is little or no significance apart from the book-to-market ratio, supporting existing literature, mutual fund share (which may be purely coincidental) and, of course, the constant.

Regression 2 performs the same type of regression yet with only one predictor, our market fragility beta. The associated coefficient is not significant, further supporting that there is no relationship between these variables depicted in figure 3.10 above. Yet, a thorough investigation is required in order to infer this result.

In Regression 3, we regress market return betas against market fragility betas while controlling for endogeneity as well as ensuring robust standard errors. The results are very similar to that of Regression 1 and 2. Thus, we may now infer that our market fragility betas are insignificant, for reasons discussed above. There is no more discernible information that we may discuss further here.

In Regressions 4 to 9, we apply the same methodology as we did for Regressions 1 to 3. For the most part, the results are very similar and we may thus apply the same arguments here as we did for Regressions 1 to 3. The only discernible difference is the weakening of the overall significance of the coefficients. This result is unusual considering that we down weight those stocks with low book value and market cap, whose data is more likely to contain sampling error. Thus, we can infer very little from these regressions.

Figure 3.13: Several regressions of Market, HML and SMB betas against their respective fragility betas as well as other controls

	Market Return Beta: β_{it+1}			HML Return Beta: β^{HML}_{it+1}			SMB Return Beta: β^{SMB}_{it+1}		
Predictors/Regressions	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	0.09389	0.08494	0.09920	0.07843	0.10754	0.09390	0.11392	0.11839	0.10920
	3.42910	3.42924	3.38303	3.27290	2.89034	3.29503	2.29431	2.55803	2.17499
G^{β}_{it}		9.42496	1.34839						
		0.68394	0.85903						
G^{HML}_{it}					8.27321	1.27942			
					0.57389	0.68804			
G^{SMB}_{it}								8.94204	1.47392
								0.57394	0.76939
Log(# owners)$_t$	0.12821		0.12743	0.11140		0.10304	13.40400		12.24892
	1.77482		2.05730	1.51304		2.01849	1.56883		1.93982
(mf)$_t$	1.36882		1.27429	1.39429		1.13782	1.42774		1.40829
	2.32412		2.27294	1.84021		1.91414	2.13080		2.29249
log(Firm Size/100)	1.53023		1.52479	1.63934		1.61953	1.44984		1.42382
	0.42409		0.37920	0.46283		0.43839	0.41113		0.38030
log(Market Cap/100)	1.62773		1.59232	1.66482		1.56042	1.74250		1.69918
	0.32948		0.27283	0.32942		0.22948	0.37295		0.28025
Book/Market	0.95305		0.93203	0.88240		0.79298	0.87293		0.82739
	3.20428		2.56502	3.79284		3.65033	4.10310		3.92048
R^2	0.43	0.08	0.45	0.37	0.07	0.4	0.31	0.09	0.33
Regression Technique	FM	FM	FM (NW)	FM	FM	FM (NW)	FM	FM	FM (NW)
Weighting	EW	EW	EW	EW	EW	EW	EW	EW	EW

3.4 Fragility & Arbitrage

In the same format as we described the results of the regressions regarding fragility above, in table 3.14 we provide the results concerning the prediction of the one-step ahead volatility of returns against estimates of fragility, adjusted to test how arbitrageurs respond to flow induced trading. Referring to the methodology in 2.5 above, we do not provide the results of the first regression. In column 1 ("Regression 1") we perform a Fama-Macbeth regression, where observations at each cross section are equally weighted, and volatility is regressed against fragility, the sensitivity of total return volatility to stock price fragility ($|1 + \gamma_{it}|$) and their associated interaction term. The results from Regression 1 indicate that there is little or no significance apart from $|1 + \gamma_{it}|$. This coefficient is below 1 suggesting that volatility is less sensitive to fragility because arbitrageurs act to mitigate this effect.

The results from Regression 2 and 3 are similar to that of Regression 1, further providing support to the above mentioned suggestion. Overall, despite the lack of significance in the coefficient of fragility (most probably resulting in the lack of significance in the interaction term), it seems that arbitrageurs do respond to flow-induced trading based on the coefficient concerning the sensitivity of total return volatility to stock price fragility. We note that our results are less significant to those in Greenwood and Thesmar (2011), which suggests that arbitrageurs have a different impact on flow-induced trading in emerging markets. These differences could be driven by lower liquidity and ownership concentrations which restrict the ability to exploit arbitrage opportunities in emerging markets.

Figure 3.14: Several regressions of total volatility against fragility and active buys

	Total Return Volatility: σ_{it+1}		
Predictors/Regressions	(1)	(2)	(3)
Constant	0.08392	0.89242	0.08392
	9.12081	8.13810	9.01832
$\sqrt{G_{it}}$	2.18289	2.37829	2.18289
	1.83847	1.84278	1.90320
$\|1+\gamma_{it}\|\sqrt{G_{it}}$	1.28920	1.49239	1.28920
	1.63875	1.64234	1.67357
$\|1+\gamma_{it}\|$	0.75939	0.67903	0.75939
	2.23994	2.48248	2.37847
R^2	0.10	0.11	0.10
Regression Technique	FM	FM	FM (NW)
Weighting	EW	MF	EW

4

Conclusions

This research project replicated and built on a paper by Greenwood and Thesmar (2011). The authors constructed a measure, expressed as a function of an asset's ownership structure and the covariance matrix of flows into and out of mutual funds, known as stock price fragility. The authors applied the measure by investigating it as a proxy for non-fundamental risk, serving as a predictor of total return volatility. They extended this measure with the construction of co-fragility, scaled co-fragility and the fragility beta and investigate them as predictors of the covariances, correlations and betas of stock returns respectively. The authors also extended the measure to investigate the sensitivity of stock price fragility to total return volatility. In this research project, we applied the same principles, but in the context of an emerging market, South Africa. Moreover, we attempted to account for potential heteroskedasticity and autocorrelation when constructing the prediction models.

Greenwood and Thesmar (2011) were able to empirically justify stock price fragility and its associated extensions using mutual funds in the US market. In this research project, the aim was to add further justification to the measure as being a proxy for non-fundamental risk using available South African funds as well as to explain the effect of extending the measure in this developing market. Moreover, it was hypothesized that the measure and its extensions hold greater tractability in an emerging market context. Unfortunately, we did not succeed in obtaining sufficient data to provide support to the results illustrated in Greenwood and Thesmar (2011). Nevertheless, with data complications and sampling errors inherent in the modelling, we were able to adequately investigate the measure and its extensions in the context of the South African market such that some interesting results were produced. Firstly, stock price fragility proved to be a significant predictor of total return volatility. Given flow data complications, we may have only attributed this result to the ownership structure of funds. This notion was supported by the significance in mutual fund share, the number of funds owning the stock as well as ownership concentration. Secondly, although co-fragility and the fragility beta did not prove to be significant against the covariance and beta of returns respectively, the regressions gave some indication of the significance in the ownership structure components of these extensions. Thirdly, through the active buys within the funds, we did show that arbitrageurs affect the sensitivity of stock price fragility to total return volatility, specifically by dampening it. Fourthly, we supported existing literature by illustrating the overall significance of the fundamental controls that predict total return volatility. Finally, although we did not show whether heteroskedasticity and autocorrelation exists in the models, by implementing the Newey-West estimator (1987), if these issues did exist, we obtained robust standard errors with very little change in the significance of the coefficients.

5

References

Anton, M., Polk, C., 2010. *Connected stocks.* Working Paper, London School of Economics.

Bae, K. H., Chan, K., Ng, A., 2004. Investibility and return volatility. *Journal of Financial Economics*, 71(2), pp. 239-263.

Basiewicz, P., Auret, C., 2010. Feasibility of the Fama and French three factor model in explaining returns on the JSE. *Investment Analysts Journal*, 71, pp. 13-25.

Bekaert, G., Harvey, C., Lundblad, C., 2007. Liquidity and expected returns: Lessons from emerging markets. *Review of Financial Studies*, 20(6), pp. 1783-1831.

Brunnermeier, M., Nagel, S., 2005. Hedge funds and the technology bubble. *The Journal of Finance*, 59(5), pp. 2013-2040.

Caicedo-Llano, J., Dionysopoulos, T., 2007. Predictors of Stock Returns in Emerging Equity Markets. *LACEA and LAMES Joint Annual Meeting.* Bogota, Colombia 4-6 October 2007.

Carhart, M., 1997. On Persistence in Mutual Fund Performance. *Journal of Finance*, 52(1), pp. 5782.

Cont, R., Wagalath, L., 2012. *Fire sales forensics: measuring endogenous risk.* Working paper.

Coval, J., Stafford, E., 2007. Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, 86, pp. 479-512.

Danielsson, J., Shin, H.S., 2003. *Endogenous Risk in Modern Risk Management: A His-*

tory. Risk Books.

Fama, E., French, K., 1992. The Cross-Section of Expected Stock Returns. *Journal of Finance*, 47(2), pp. 427-465.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, pp. 3-56.

Fama, E., Macbeth J., 1973. Risk, Return, and Equilibrium: Empirical Tests. *The Journal of Political Economy*, 81(3), pp. 607-636.

Frazzini, A., Lamont, O. A., 2008. Dumb money: mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics*, 88, pp. 299-322.

Gormley, T., Matsa, D., 2012. *Common errors: How to (and not to) control for unobserved heterogeneity*. Working Paper, University of Pennsylvania.

Greenwood, R., Thesmar, D., 2011. Stock price fragility. *Journal of Financial Economics*, 102(3), pp. 471-490.

Hirschman, A. O., 1964. The paternity of an index. *The American Economic Review*, 54(5), pp. 761-762.

Jefferis, K., Smith, G., 2005. The changing efficiency of African stock markets. *South African Journal of Economics*, 73, pp. 54-67.

Jotikasthira, C., Lundblad, C., Ramadorai, T., 2009. *Asset fire sales and purchases and the international transmission of financial shocks*, CEPR Discussion Papers.

Kaminsky, G., Lyons, R., Schmukler, S., 2001. Mutual fund investment in emerging markets: An overview. *The World Bank Economic Review*, 15(2), pp. 315-340.

Koch, A., Ruenzi, S., Starks, L., 2010. *Commonality in liquidity: a demand-side explanation*. Working Paper, University of Texas.

Kurz, M., Motolese, M., 2001. Endogenous uncertainty and market volatility. *Economic Theory*, 17(3), pp. 497-544.

Lin, C., 2011. *Investor Sentiment and the Fragility of Liquidity*. Working Paper.

Lin, C., Massa, M., Zhang, H., 2011. *Stock Market Fragility and the Quality of Governance of the Country*. Working Paper.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47 (1), pp. 13-37.

Lou, D., 2010. *A flow-based explanation for return predictability*. Working Paper, London School of Economics.

Mangani, R., 2008. Modelling Return Volatility on the JSE Securities Exchange of South Africa. *The African Finance Journal*, 10(1), pp. 55-70.

Markowitz, H., 1952. Portfolio selection. *Journal of Finance*, 7(2), pp. 77-91.

Mhlanga, G., 2008. *The covariation of South African and foreign equity returns during bull and bear runs: Implications for portfolio diversification*. Ph. D. , Rhodes University.

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation-consistent covariance matrix. *Econometrica*, 55, pp. 703-708.

Pillay, N., Muller, C., Ward, M., 2010. Fund size and returns on the JSE. *Investment Analysts Journal*, 71, pp. 1-11.

Qin, Y., 2007. Liquidity and commonality in emerging markets. In *20th Australasian Finance & Banking Conference*. Sydney, Australia 12-14 December 2007.

Ross, S., 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3), pp. 341-360.

Samouilhan, N. L., Shannon, G., 2008. Forecasting Volatility on the Johannesburg Stock Exchange. *Investment Analysts Journal*, 67, pp. 19-28.

Sharpe, W., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19 (3), pp. 425-442.

Treynor, J., 1962. *Toward a Theory of Market Value of Risky Assets*. Unpublished Manuscript.

Van Rensburg, P., Robertson, M., 2003. Size, Price-to-Earnings and Beta on the JSE. *Investment Analysts Journal*, 58, pp. 1-11.

Yartey, C. A., 2008. *The Determinants of Stock Market Development in Emerging Economies: Is South Africa Different* IMF Working Paper 08/38, Washington DC: International Monetary Fund.

Appendix A

Stock and Industry Data

The following example has been taken from: http://www.cipc.co.za/SIC.aspx

"The SIC code consists of a 5 digit number with each digit of the code having the following significance:

First Digit = Major Division Second Digit = Division Third Digit = Major Group Fourth Digit = Group Fifth Digit = Sub-Group

If, as an example, we look at the SIC code 33711, we will be able to extract the following meaning:
3: The first digit or Major Division = Manufacturing.
3: The second digit or Division = Manufacture of coke, refined petroleum products and nuclear fuel; manufacture of chemicals and chemical products; manufacture of rubber and plastic products.
7: The third digit or Major Group = Manufacture of rubber products.
1: The fourth digit or Group = Manufacture of rubber tyres and tubes; retreading and rebuilding of rubber tyres.
1: The fifth digit or Sub-Group = Manufacture of tyres and tubes."

For this research project, we are only concerned with the first four digits related to each stock. The SIC code for each stock can be seen in the table below:

Share Name	Share Code	SIC Code
African Bank Investments Ltd	ABL	8773
Arcelormittal SA Ltd	ACL	1757
Aveng Ltd	AEG	2357
Anglo American PLC	AGL	1775
Anglo American Platinum Ltd	AMS	1779
Anglogold Ashanti Ltd	ANG	1777
Aspen Pharmacare Holdings Ltd	APN	4577
African Rainbow Minerals Ltd	ARI	1775
Absa Group Ltd	ASA	8355
Assore Ltd	ASR	1775
Barloworld Ltd	BAW	2727
BHP Billiton Ltd & Gas	BIL	1775
The Bidvest Group Ltd	BVT	2727
Capital Shopping Centres Group PLC	CSO	8637
Exxaro Resources Ltd	EXX	1771
Firstrand Ltd	FSR	8355
Gold Fields Ltd	GFI	1777
Growthpoint Properties Ltd	GRT	8633
Harmony Gold Mining Company Ltd	HAR	1777
Impala Platinum Holdings Ltd	IMP	1779
Investec Ltd	INL	8777
Imperial Holdings Ltd	IPL	2777
Mr Price Group Ltd	MPC	5371
Massmart Holdings Ltd	MSM	5373
MTN Group Ltd	MTN	6575
Murray & Roberts Holdings Ltd	MUR	2357
Nedbank Group Ltd	NED	8355
Naspers Ltd	NPN	5553
Old Mutual PLC	OML	8575
Remgro Ltd	REM	2727
RMB Holdings Ltd	RMH	8355
Sabmiller PLC	SAB	3533
Sappi Ltd	SAP	1737
Standard Bank Group Ltd	SBK	8355
Steinhoff International Holdings Ltd	SHF	3726
Shoprite Holdings Ltd	SHP	5337
Sanlam Ltd	SLM	8575
Sasol Ltd	SOL	0537
Tiger Brands Ltd	TBS	3577
Truworths International Ltd	TRU	5371
Woolworths Holdings Ltd	WHL	5373

Table A.1: Table showing share names, share codes and SIC codes.

Appendix B

Some Matlab Code

```
% This function computes the fund flows as well as the variance-covariance
% matrix of these flows.
function [F covF] = flows(a,R)

% a (fund sizes) and R (fund returns) are k by t+1 since lagged fund sizes
% are used to compute flows.
[m N] = size(a);
n = N-1;

% Flows calculation.
% Assume flows must be t by k. Here Ft is k by t for convenience. We
% transpose later.
Ft = a(:,2:n+1) - a(:,1:n).*(ones(m,n) + R(:,2:n+1));

% Normalize flows.
nrmFt = Ft./a(:, 1:n);

% F must be a t by k matrix, resulting in a kxk covariance matrix.
F = Ft';
nrmF = nrmFt';

% predefine variance-covariance matrix of flows to speed computation time.
covF(1:n) = {zeros(m,m)};
% Compute variance-covariance atrix of flows.
for t = 1:n
    covF(t) = {diag(a(:,t))*cov(nrmF(1:t,:))*diag(a(:,t))};
end



end
```

```
% This function is responsible for computing stock price fragility and its
% components over all time points for the ith stock.
function [w G mf H numown ondiag offdiag] = fragility(s, p, covF, a, theta)

% assume s (matrix of the number of shares per fund and time point)
% is k (funds) by t (time).
% assume a (fund sizes) is a k by t matrix.
% assume theta (market cap) and p (share price) are row matrices 1 by t.
% covF (the variance-covariance of flows) is a 1 by t cell array containing
% k by k matrices.

% construct the weights for the ith stock in each fund per time point.
% w is a k by t matrix.
w = (s*diag(p))./a;

[m n] = size(w);

% Predefining variables to speed up computation time.
G = zeros(n);
mf = zeros(n);
D = zeros(n);
omega = zeros(n);
H = zeros(n);
numown = zeros(n);
offdiag = zeros(n);
ondiag = zeros(n);


% Here we compute fragility and its components for the ith stock.
for t = 1:n
    % Fragility
    G(t) = (1/(theta(t)*theta(t)))*(w(:,t)'*covF{t}*w(:,t));
    % Fund ownership
    mf(t) = sum(s(:,t))*p(t)/theta(t);
```

```matlab
    for k = 1:m
        % Ownership concentration
        H(t) = H(t) + (s(k,t)*p(t)/(theta(t)*mf(t)))^2;
        if s(k,t) ~= 0
            % Number of owners of the ith stock
            numown(t) = numown(t) + 1;
        end
    end
    % Diagonal elements of fragility.
    D(t) = {covF{t}.*eye(n,n)};
    % The mean of these diagonal elements.
    omega(t) = trace(D{t})/n;
    % Off-diagonal terms.
    offdiag(t) = (1/(theta(t)*theta(t)))*(w(:,t)'*(covF{t} - D{t})*w(:,t));
    % On-diagonal terms.
    ondiag(t) = (1/(omega(t)*omega(t)))*(w(:,t)'*(D{t} - diag(omega))*w(:,t)) ...
        + mf(t)*mf(t)*omega(t)*H(t);
end

end
```

```
% This function is resonsible for computing cofragility for all stocks.
function [w cG nrmcG] = cofragility(s, p, covF, a, theta, G)
% assume s (matrix of the number of shares per fund and time point)
% is a 1 by i cell array with k (funds) by t (time) matrices.
% assume a (fund sizes) is a k by t matrix.
% assume theta (market cap) and p (share price) are i by t matrices.
% covF (the variance-covariance of flows) is a 1 by t cell array containing
% k by k matrices.
[num n] = size(p);
m = size(covF{1});

% predefining variables
[w{1:num}] = deal(zeros(m,n));
[cG{1:n}] = deal(zeros(num,num));
[nrmcG{1:n}] = deal(zeros(num,num));

% construct the weights for the ith stock in each fund per time point.
% w is a 1 by i cell array containing k by t matrices.
for i = 1:num
w{i} = (s{i}*diag(p(i,:)))./a;
end

for t = 1:n
    for i = 1:num
        for j = 1:num
            % Co-fragility
            cG{t}(i,j) = (1/(theta(i,t)*theta(j,t)))*(w{i}(:,t)'*covF{t}*w{j}(:,t));
            % Normalized co-fragility for to test against return
            % correlations.
            nrmcG{t}(i,j) = cG{t}(i,j)/sqrt(G(i,t)*G(j,t));
        end
    end
end

end
```

```
% This function is responsible for constructing the fragility beta.
function FBeta = fragilitybeta(cG, pw)

% cG is co-fragility: a 1 by t cell array containing i by i matrices.
% pw will be a row vector comprising of portfolio weights for each stock.

n = size(cG,2);
num = size(cG{1},1);

% Predefine variables to speed up computation time.
[covFpFi{1:n}] = deal(zeros(num));
varFp = zeros(n);
FBeta = zeros(num,n);


for t = 1:n
    for i = 1:num
        for j = 1:num
            % Numerator.
        covFpFi{t}(i) = covFpFi{t}(i) + pw(j)*cG{t}(i,j);
        end
        % Denominator.
        varFp(t)= varFp(t) + pw(i)*pw(i)*cG{t}(i,i);
    end
end

% Fragility Beta Consttruction (an i by t matrix).
for t = 1:n
    for i = 1:num
        FBeta(i,t) = covFpFi{t}(i)/varFp(t);
    end
end


end
```

```
% This function is responsible for constructing the order imbalances driven
% by active trades.
function [flows activetrades] = arbitrage(s, w, p, theta, F)
num = size(p,1);
[m n] = size(w{1});
F = F';

% predefine variables
[flows{1:num}] = deal(zeros(m,n));
% Weighted flows: a 1 by t cell array containing k by t matrices.
% These weighted flows are used as the independent variable in the
% first-step regression of order imbalances.
for i = 1:num
    flows{i} = (w{i}.*F)*diag(1./theta(i,:));
end

dp(:,1) = 0;
dp(:,2:n) = p(:,1:n-1) - p(:,2:n);
sumwp = zeros(m,n);
for i = 1:num
    sumwp = sumwp + w{i}*diag(dp./p);
end

[activetrades{1:num}] = deal(zeros(m,n));
[dw{1:num}] = deal(zeros(m,n));
% The order imbalance driven by active trades are computed and are used as
% the dependent variable in the first-step regression.
for i = 1:num
    dw{i}(:,1) = 0;
    dw{i}(:,2:n) = w{i}(:,1:n-1) - w{i}(:,2:n);
    activetrades{i} = s{i}*diag(p)*(dw{i}./w{i} - (dp./p - sumwp))*diag(1./theta(i,:));
end

end
```

```
function [b se t] = FamaMacBeth_NW(y, X, PART_VAR, beta, varargin)
% Routine for calculating Newey-West-adjusted Fama-MacBeth standard errors.
% SYNTAX: ret = FamaMacBeth_NW(y, X, PART_VAR, BETA, VARARGIN)
%   y: vector of dependent variable
%   X: matrix of regressors
%   PART_VAR: vector containing variable by which regressions are
%     partitioned  (e.g. MONTH, using the Fama-MacBeth approach on
%     a data set in which MONTH uniquely identifies a month)
%   BETA: The vector of coefficients under the null hypothesis.
%   VARARGIN: Leave empty to get unadjusted Fama-MacBeth estimates.
%     Use 'NW' followed by the lag length for Newey West (1987)
%     correction
%     LAG_LENGTH: The number of lags to be considered in the Newey-West
%     approach.
%   RET = [b se t], the estimated coefficients (b), the estimated
%     standard errors (se), and t-statistics (t).

  k = size(X,2);
  % Get details on partition variable.
  PART = unique(PART_VAR);
  T = length(PART);
  % Create a location for estimated coefficients.
  temp = zeros(k, T, 4);
  % Get Fama-MacBeth standard error
  % First estimate t coefficients, one for each year
  bb=zeros(T,k);

  for t=1:T
    % Get y and X values for regression t of T
    y2 = y(find(PART_VAR==PART(t)),:);
    X2 = X(find(PART_VAR==PART(t)),:);
    % Store estimated coefficient, SEs, t-statistics, and df
    temp(:,t,1:3) = regress(y2,X2,0);
    temp(:,t,4) = size(X2,1)-size(X2,2);
```

```
    % Store estimated coefficients for t in BB vector
    bb(t,:) = temp(:,t,1);
  end

  % Calculate standard errors using the time-series distribution of the
  % estimated coefficients.
  b = zeros(1:k);
  se = zeros(1:k);
  for i=1:k
    b(i) = mean(temp(i,:,1));
    se(i) = 1/sqrt(T) * std(temp(i,:,1));
  end

  % Calculate associated t-statistics.
  if nargin ~= 4
    if nargin < 4 || nargin > 6
    error('Wrong number of arguments');
    end
    if varargin{1} =='NW'
      % Calculate Newey-West standard errors.
      lag_length = varargin{2};
      for i=1:k
        oness=ones(T,1);
        ret=NeweyWestPanelStata(bb(:,i), oness, lag_length, oness, oness, 0);
        se(i)=ret(1,2);
      end
  end

  t = (b-beta') ./ se;
  b=b';
  se = se';
  t = t';

end
```

```
function ret = NeweyWestPanelStata(y, X, L, FIRM_VAR, TIME_VAR, stat)
% Function that runs OLS and returns standard errors based
% Newey-West autocorrelation consistent covariance estimator.
% This routine can be used for a single time-series or panel data.
% SYNTAX: ret = NeweyWestPanelStata(y, X, L, FIRM_VAR, TIME_VAR, stat)
% The value returned depends on stat:
%  0: [beta, standard errors, t-statistics]
%  1: residuals

  % Sort the data by FIRM_VAR and then TIME_VAR
  FIRMS = unique(FIRM_VAR);
  j = 0;
  for i=1:length(FIRMS);
    rows = find(FIRM_VAR==FIRMS(i));
    [temp, ix] = sort(TIME_VAR(rows,:));
    rows = rows(ix,:);
    if i==1
      y2 = y(rows,:);
      X2 = X(rows,:);
      F2 = FIRM_VAR(rows,:);
      T2 = TIME_VAR(rows,:);
    else
      y2 = [y2; y(rows,:)];
      X2 = [X2; X(rows,:)];
      F2 = [F2; FIRM_VAR(rows,:)];
      T2 = [T2; TIME_VAR(rows,:)];
    end
  end
  % Reassign the variables after sorting.
  y = y2;
  X = X2;
  FIRM_VAR = F2;
  TIME_VAR = T2;
  % calculate Bhat
  b = pinv(X)*y;
```

```
[N, k] =size(X);
% Generate residuals
e = y - X * b;

% Calculate the Newey-West autocorrelation consistent covariance
% estimator.
Q = 0;
for l = 0:L
  w_l = 1-l/(L+1);
  for t = l+1:N
    if (l==0)    % This calculates the S_0 portion
      Q = Q  + e(t) ^2 * X(t, :)' * X(t,:);
    else          % This calculates the off-diagonal terms
      if FIRM_VAR(t,1) == FIRM_VAR(t-1,1)
      Q = Q + w_l * e(t) * e(t-l)* ...
        (X(t, :)' * X(t-l,:) + X(t-l, :)' * X(t,:));
      end
    end
  end
end
Q = 1/(N-k) * Q;

% Calculate Newey-West standard errors
varBhat = N * inv(X' * X) * Q * inv(X' * X);
% calculate standard errors and t-stats
se = sqrt(diag(varBhat));
t = b ./ se;
if (stat==1) % return residuals
  ret = e;
elseif (stat==0) % return SSR
  ret = [b se t];
end
end
```

```
function results = pfixed(y,index,x)
% PURPOSE: performs Fixed Effects Estimation for Panel Data
%           (for balanced or unbalanced data)using the within-groups
%              estimation procedure.
% USAGE:  results = pfixed(y,index,x)
% where: y: This matrix must include in the firt column the dependent
%         variable, the independent variables must follow accordingly.
%         index: index vector that identifies each observation with an individual
%                    e.g. 1  (first 2 observations  for individual # 1)
%                         1
%                         2  (next  1 observation   for individual # 2)
%                         3  (next  3 observations  for individual # 3)
%                         3
%                         3
%         x: optional matrix of exogenous variables, dummy variables.
% RETURNS a structure:
% results.beta  = bhat
% results.tstat = t-statistics
% results.tprob = t-probabilities
% results.rsqr  = r-squared

[nobs equ]= size(y);
nx = 0;
if nargin == 3
[nobs2 nx] = size(x);
 if (nobs2 ~= nobs)
 error('nobs in x-matrix not the same as y-matrix');
 end;
end;

% creation of the id matrix using the vector index
nindiv = length(unique(index));
id = zeros(nindiv,3);
id(:,1) = unique(index);
```

```
for i=1:nindiv
    id(i,2) = length(find(index == i));
end;
id(:,3) = cumsum(id(:,2));

% tranformation of all the variables used
% the variables are expressed as deviations from the individual means
[n u]= size(id);
i=1;
for j=1:n
    while i<=id(j,3),
      ytemp= y(i:id(j,3),:);
      medias(j:id(j,1),:)= mean(ytemp);
      madj(i:id(j,3),:)= ytemp-(ones(id(j,2),1)*mean(ytemp));
      i= i+ id(j,2);
    end;
end;
y= madj;

% form x-matrix
if nx
xmat = [y(:,2:equ) x];
else
xmat = [y(:,2:equ)];
end;
[nobs3 nvars]= size(xmat);
% run OLS
 res = ols(y(:,1),xmat);
 results.beta  = res.beta;      % bhats
 results.tstat = res.tstat;     % t-stats
% compute t-probs
      tstat = zeros(nvars,1);
      tstat = res.tstat;
      tout = tdis_prb(tstat,nobs-nvars);
 results.tprob = tout;          % t-probs
   results.rsqr = res.rsqr;      % r squared
```